

Two-stage Geometric Information Guided Image Reconstruction

Jing Qin *

Weihong Guo †

Abstract

In compressive sensing, it is challenging to reconstruct image of high quality from very few noisy linear projections. Existing methods mostly work well on piecewise constant images but not so well on piecewise smooth images such as natural images, medical images that contain a lot of details. We propose a two-stage method called GeoCS to recover images with rich geometric information from very limited amount of noisy measurements. The method adopts the shearlet transform that is mathematically proven to be optimal in sparsely representing images containing anisotropic features such as edges, corners, spikes etc. It also uses the weighted total variation (TV) sparsity with spatially variant weights to preserve sharp edges but to reduce the staircase effects of TV. Geometric information extracted from the results of stage I serves as an initial prior for stage II which alternates image reconstruction and geometric information update in a mutually beneficial way. GeoCS has been tested on incomplete spectral Fourier samples. It is applicable to other types of measurements as well. Experimental results on various complicated images show that GeoCS is efficient and generates high-quality images.

Keywords: compressive sensing, shearlet transform, weighted TV, split Bregman, ADMM

1 Introduction

1.1 Background

Compressive sensing (CS) (cf. the pioneering work [1, 2]) is a new technique in data acquisition realm that reconstructs the signal from fewer measurements than that required by the classical Nyquist-Shannon sampling theorem. This fact makes it very useful in reducing the sensing cost in a variety of applications such as geophysics, astronomy and medical imaging.

The incomplete measurement $b \in \mathbb{C}^m$ of CS is usually linear projections of the underlying image of interest $\bar{u} \in \mathbb{C}^n$ in the form of $b = A\bar{u} + r$ where r is noise. It relies on sparsity/compressibility of \bar{u} itself or under certain transform Φ to recover it from b . When there is no noise ($r = 0$), a straightforward approach to reconstructing \bar{u} is to solve the L_0 problem

$$\min_u \|\Phi u\|_0 \quad \text{s.t.} \quad Au = b. \quad (1)$$

However, since the L_0 problem is NP-hard, it is common in practice to consider a convex relaxed L_1 optimization problem

$$\min_u \|\Phi u\|_1 \quad \text{s.t.} \quad Au = b, \quad (2)$$

which is more computationally efficient. When the sparsifying transform Φ is orthonormal, the solution of (2) turns out to be the same as that of (1), and approximate the underlying signal with an overwhelming probability if A satisfies the restricted isometry property (RIP). A commonly used example of orthonormal Φ is the Haar wavelet transform. More recently, RIP is generalized to D-RIP [3], a property that guarantees accurate recovery of images that are nearly sparse in overcomplete/redundant dictionaries. The theoretical results make it more flexible to choose Φ and reconstruct the signal using L_1 optimization (2). In the presence of noise, a relaxed model of interest is

$$\min_u \|\Phi u\|_1 \quad \text{s.t.} \quad \|Au - b\|_2 \leq \sigma,$$

*Department of Mathematics, University of California, Los Angeles, CA 90095, USA. Email: jxq@ucla.edu

†Department of Mathematics, Case Western Reserve University, Cleveland, OH 44106, USA. Email: wxg49@case.edu

or equivalently

$$\min_u \lambda \|\Phi u\|_1 + \frac{1}{2} \|Au - b\|_2^2,$$

where $\lambda > 0$ is related to the noise level σ in the data. To preserve the discontinuities of u which correspond to the image features, e.g., edges, total variation (TV) is taken into consideration as an additional regularization term [4, 5, 6]. Then the model with two regularization terms reads as

$$\min_u \beta \|u\|_{\text{TV}} + \lambda \|\Phi u\|_1 + \frac{1}{2} \|Au - b\|_2^2 \quad (3)$$

where Φ is the wavelet transform and β depends on the characteristics of the underlying image itself.

Wavelet transform and total variation have been used widely in various computer vision and/or imaging science problems. The advantage of wavelet transform is its optimality in approximating signals containing point-wise singularities, but it is widely known that traditional wavelets are not so effective in dealing with singularities in higher dimensions, such as edges in 2D images. TV is optimal in describing piecewise constant images and preserving image edges. However, it is well known that the TV regularization will cause staircase effects. Recently, it has been shown that the TV regularization is closely connected to the wavelet one [7]. Therefore, the combination of wavelet and TV is not ideal for reconstructing natural images with abundant directional geometric information from few noisy CS measurements.

In this paper, we present a geometric information guided CS (abbreviated as GeoCS) reconstruction method to improve the performance for the situations when the sampling rate is low and/or the noise level is high. The goal is to preserve geometries and fine features with less data required than the state-of-the-art methods. There are three major contributions in our paper.

- Shearlet transform instead of the widely used wavelet transform is adopted as Φ to promote the sparsity of signals and thereby reduce the number of measurements required for accurate recovery of them by the fundamental CS theory. Shearlets [8, 9, 10] provide an optimally sparse approximation of piecewise smooth function with C^2 singularity curves, e.g., edges, cusps and corners. It combines the power of multiscale methods with the capability of extracting geometry of images.
- A two-stage method rather than the conventional one-stage method is applied to obtain better recovered images from fewer measurements than the state of the arts. This reconstruction approach adaptively learns the gradient of increasing accuracy to some extent controlling the image geometry. The first stage is to get an initial image reconstruction from CS measurements based on shearlet transform and TV. The second stage starts with geometric information extracted from the result of stage I, and alternates image reconstruction and geometric information update until it converges. Adaptive TV and shearlet transform are used in the second stage. Massive numerical experiments show that the two-stage approach outperforms the classical one-stage methods.
- Apply the alternating direction method of multipliers (ADMM) [12, 13] or its equivalent split Bregman method [11] to efficiently solve the optimization problems in both stages. The algorithm achieves fast convergence and produces high-quality images. The convergence of the algorithm at each stage is guaranteed as well.

In [14], the binary reweighted L_1 regularization is exploited for recovering 1D sparse signals. While it can be applied to recover the sparse wavelet coefficients of an image and hence the image itself, it can hardly take advantages of image edges to improve the recovery. In our recent work [15, 16], binary edge detection and image reconstruction are performed alternatively in a mutually beneficial way and thus the recovery quality has been improved. This work considers the more general edge detection whose range is continuous rather than binary. The spatially variant weights associated to the TV ranges between zero and one based on the extracted salient geometric information. In that case, the sharp edges are still able to be preserved while gradual intensity changes in smooth regions can be preserved as well to reduce the staircase effects resulted in TV regularization.

The paper is organized as follows. We provide a brief review of the shearlet transform in Section 2, and present the two-stage geometric information guided algorithm in Section 3. The convergence analysis of the algorithm and practical parameter selection are presented in Section 4. To show the consistent excellence

and robustness of the proposed algorithm, plenty of numerical results and comparisons to related work reconstruction from partial Fourier data (RecPF) [17] and Edge Guided Compressive Sensing (EdgeCS) [16] are provided in Section 5. Finally, conclusion and remarks are made in Section 6.

2 Review of shearlet transform

The traditional wavelet transform is based on isotropic dilations and thus has limited ability to describe the geometry of multidimensional functions. Directional representation systems such as ridgelets [18], curvelets [19], contourlets [20], and shearlets [9] have been designed to provide much more geometric information of multidimensional functions such as images. Curvelets, a tight frame of elongated oscillatory functions at various scales, was first proposed by Candès and Donoho to generalize wavelet. For any $L^2(\mathbb{R}^2)$ function f , the N largest term approximation using the curvelet transform has error norm of order $(\log N)^3 N^{-2}$. Since the curvelets are not generated by taking a family of actions on one function as wavelet, it is numerically difficult to implement. In an attempt to provide a better discrete implementation of the curvelets, the contourlet representation is then proposed. It is a discrete time-domain construction, which is designed to achieve essentially the same frequency tiling as the curvelet representation. With the same rate of approximation error decay as curvelets, shearlets have several advantages: efficient implementation, more directional sensitivity and theoretical relation to the multiresolution analysis.

Shearlet transform is an efficient multiscale directional representation of signals, theoretically proven to be optimal up to a log-factor in encoding images with anisotropic features such as edges, corners and other singularities [8, 10]. Given any function $\psi \in L^2(\mathbb{R}^2)$, the shearlet system is generated by applying the operations of dilation, shear transformation and translation of ψ :

$$\psi_{ast} = |\det M_{as}|^{-\frac{1}{2}} \psi(M_{as}^{-1}(x - t))$$

where

$$M_{as} = \begin{bmatrix} a & -\sqrt{as} \\ 0 & \sqrt{a} \end{bmatrix} = \begin{bmatrix} 1 & -s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & \sqrt{a} \end{bmatrix} := B_s A_a$$

with $a \in \mathbb{R}^+$, $s \in \mathbb{R}$, $t \in \mathbb{R}^2$, B_s a shear operator and A_a an anisotropic dilation operator. Note in the generation of wavelets, there are only isotropic scaling and translation involved without shearing or anisotropic scaling. The shearlet transform of function $f \in L^2(\mathbb{R}^2)$ is defined as

$$SH_\psi(f)(a, s, t) = \langle f, \psi_{ast} \rangle.$$

The shearlet transform is invertible if the function ψ satisfies the admissibility property

$$\int_{\mathbb{R}^2} \frac{|\widehat{\psi}(\omega_1, \omega_2)|^2}{|\omega_1|^2} d\omega_1 d\omega_2 < \infty$$

where $\widehat{\psi}$ is Fourier transform of ψ . Given complete shearlet transform coefficients, the original function f can be recovered by

$$f(x) = \int_{\mathbb{R}^2} \int_{\mathbb{R}} \int_{\mathbb{R}^+} \langle f, \psi_{ast} \rangle \psi_{ast}(x) \frac{da}{a^3} ds dt.$$

Discrete shearlet transform can be implemented efficiently using the fast Fourier transform. There are three shearlet toolboxes using MATLAB available online: Local Shearlet Toolbox <http://www.math.uh.edu/~dlabate/software.html>, Fast Finite Shearlet Transform (FFST)[21], and ShearLab <http://www.shearlet.org>. More recently, the shearlet transform has been successfully applied in image processing, e.g., the shearlet-based total variation denoising algorithm [22]. Because of its higher sparsity of signal representation and ability to capture directional features, the performance of the shearlet transform has also been explored in CS field [23].

3 Proposed model and algorithm

In this section, we present our reconstruction model and analyze how to apply split ADMM to solve the model at each stage. The idea is to use both the shearlet transform and the weighted TV. To enhance the accuracy of weights associated to the TV regularization, we propose a two-stage method. The first stage is to solve a standard TV- L_1 - L_2 model with the shearlet transform to get an initial guess for the underlying image of interest. Note that the extraction of geometry does not work at Stage I since the accuracy is relatively low. In the second stage, we generate the initial spatially variant weights based on the result from stage I, and then alternate image reconstruction and weights update until it converges at this stage. The entire algorithm alternates the two stages until the relative error between two consecutive results is within a tolerance value.

For the shearlet part Φu , we adopt the FFST algorithm which is completely done by the Fourier transform and the inverse Fourier transform. Let

$$\Phi u := SH(u) = \sum_{i=1}^N SH_i(u)$$

where $SH_i(u)$ is the i th subband of shearlet transform of u and N depends on the number of scales in shearlet transform. The i th subband of the shearlet transform can be efficiently implemented as componentwise multiplication with a mask matrix denoted by H_i in the frequency space. We have

$$SH_i(u) = \text{vec}(\mathcal{F}^{-1}(H_i * \hat{U})) = \text{vec}(\mathcal{F}^{-1}(H_i) * U) := M_{H_i}u$$

where $M_{H_i} \in \mathbb{R}^{n^2 \times n^2}$, U is the matrix representation of the vectorized image u and \hat{U} is the Fourier transform of U .

We demonstrate the idea of the proposed model using partial Fourier sampling, but it can be extended easily to other linear projection measurements. Let $A = \mathcal{F}_p := PF$ where P is a selection matrix and F is the Fourier transform operator. For 2D Fourier transform, $F \in \mathbb{R}^{n^2 \times n^2}$ is the Kronecker product of two identical $n \times n$ unitary Fourier transform matrices G with

$$G_{jk} = \frac{1}{\sqrt{n}} e^{-2\pi\sqrt{-1}(k-1)(j-1)/n}, \quad j, k = 1, \dots, n.$$

It can be shown that F satisfies $F^*F = FF^* = I_{n^2}$. By this notation, we get the explicit representation of M_{H_i} as

$$M_{H_i} = F^* \text{diag}(\text{vec}(H_i)) F.$$

The selection matrix $P \in \mathbb{R}^{k \times n^2}$ is generated simply by deleting the $(n(j-1) + i)$ th row of the $n^2 \times n^2$ identity matrix if the (i, j) th entry of data matrix is not sampled.

3.1 Stage I: TV- L_1 - L_2 model

To simplify our discussion, we assume the image to be studied has a square domain. Let $u \in \mathbb{R}^{n^2}$ be the vectorized ground truth image, and $b \in \mathbb{R}^k$ ($k \ll n^2$) the given data. At the first stage, we consider the unconstrained minimization problem with anisotropic discretization of TV as follows:

$$\min_{u \in \mathbb{R}^{n^2}} \beta \sum_{i=1}^2 \|D_i u\|_1 + \lambda \sum_{i=1}^N \|SH_i(u)\|_1 + \frac{1}{2} \|\mathcal{F}_p(u) - b\|_2^2 \quad (4)$$

where $D_1 \in \mathbb{R}^{n^2 \times n^2}$ (D_2) is a horizontal (vertical) first order finite difference operator with periodic boundary conditions.

Due to the non-differentiability of both TV and L_1 terms, we introduce auxiliary variables $r_i \in \mathbb{R}^{n^2}$ ($i = 1, 2$) and $s_i \in \mathbb{R}^{n^2}$ ($i = 1, \dots, N$) such that $r_i = D_i u$ ($i = 1, 2$) and $s_i = SH_i(u)$ ($i = 1, \dots, N$) to split the variables. We wish to solve the problem

$$\min_{u, r_i, s_i} \beta \sum_{i=1}^2 \|r_i\|_1 + \lambda \sum_{i=1}^N \|s_i\|_1 + \frac{1}{2} \|\mathcal{F}_p(u) - b\|_2^2 \text{ s.t. } r_i = D_i u, \quad s_i = SH_i(u). \quad (5)$$

After adding the quadratic penalty terms, we get the following unconstrained problem:

$$\min_{u, r_i, s_i} \beta \sum_{i=1}^2 (\|r_i\|_1 + \frac{\mu}{2} \|r_i - D_i u\|_2^2) + \lambda \sum_{i=1}^N (\|s_i\|_1 + \frac{\tau}{2} \|s_i - SH_i(u)\|_2^2) + \frac{1}{2} \|\mathcal{F}_p(u) - b\|_2^2.$$

The above optimization problem is equivalent to (4) when $\mu, \tau > 0$ go to infinity. Solving (6) using the continuation scheme [24] is a straightforward method, but it is slow and leads to the ill conditioning of the problem when μ, τ are sufficiently large. We hereby apply the split Bregman, which provides fast convergence while values of μ, τ can be fixed. The split Bregman formulation is

$$\min_{u, r_i, s_i} \beta \sum_{i=1}^2 (\|r_i\|_1 + \frac{\mu}{2} \|r_i - D_i u - v_i\|_2^2) + \frac{1}{2} \|\mathcal{F}_p(u) - b\|_2^2 + \lambda \sum_{i=1}^N (\|s_i\|_1 + \frac{\tau}{2} \|s_i - SH_i(u) - t_i\|_2^2)$$

where v_i 's, t_i 's are updated by Bregman iterations

$$\begin{cases} v_i \leftarrow v_i + \gamma(D_i u - r_i), & i = 1, 2 \\ t_i \leftarrow t_i + \gamma(SH_i(u) - s_i), & i = 1, \dots, N, \end{cases}$$

with $\gamma > 0$ a parameter to be discussed later.

We finally decompose it into three sets of subproblems and apply alternating minimization scheme to get a minimizer iteratively.

$$\begin{cases} \min_{r_i} \|r_i\|_1 + \frac{\mu}{2} \|r_i - D_i u - v_i\|_2^2, & i = 1, 2 \\ \min_{s_i} \|s_i\|_1 + \frac{\tau}{2} \|s_i - SH_i(u) - t_i\|_2^2, & i = 1, \dots, N \\ \min_u \frac{\beta\mu}{2} \sum_{i=1}^2 \|r_i - D_i u - v_i\|_2^2 + \frac{1}{2} \|\mathcal{F}_p(u) - b\|_2^2 + \frac{\lambda\tau}{2} \sum_{i=1}^N \|s_i - SH_i(u) - t_i\|_2^2. \end{cases}$$

The first two subproblems are both in the form of L_1 - L_2 optimization

$$\min_{x \in \mathbb{R}^m} \delta \|x\|_1 + \frac{1}{2} \|x - v\|_2^2, \quad (\delta > 0)$$

whose solution is given by using the shrinkage operator

$$x = \text{shrink}(v, \delta) := \text{sgn}(v) .* \max\{|v| - \delta, 0\}$$

where $\text{sgn}(x) : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is componentwise sign function and $.*$ is componentwise multiplication. Then by the similar derivations, the first two subproblems have closed-form solutions using shrinkage.

To solve the last least square subproblem, we consider the corresponding normal equation

$$\beta\mu \sum_{i=1}^2 D_i^T (D_i u - r_i + v_i) + \lambda\tau \sum_{i=1}^N M_{H_i}^* (M_{H_i} u - s_i + t_i) + (PF)^* (PFu - b) = 0.$$

To circumvent the expensive computation of the inverse matrix, we multiply both sides by F and simply the solution due to the fact that F is unitary. By simplification, the solution is explicitly represented as

$$\begin{aligned} & \left(\beta\mu \sum_{i=1}^2 F(D_i^T D_i) F^* + \lambda\tau \sum_{i=1}^N F(M_{H_i}^* M_{H_i}) F^* + P^* P \right) F u \\ &= \beta\mu \sum_{i=1}^2 F(r_i - v_i) + \lambda\tau \sum_{i=1}^N F(s_i - t_i) + P^* b. \end{aligned}$$

Here $P^*P := \text{diag}(\tilde{P})$ is a $n^2 \times n^2$ diagonal matrix with diagonal value 0, 1 corresponding to nonsampled and sampled entries, respectively. Denoting $P^*b = \tilde{b}$, the solution can be further written in terms of Fourier transform

$$u = \mathcal{F}^{-1} \left((\beta\mu \sum_{i=1}^2 \mathcal{F}(r_i - v_i) + \lambda\tau \sum_{i=1}^N \mathcal{F}(s_i - t_i) + \tilde{b}) \right. \\ \left. ./ (\beta\mu \sum_{i=1}^2 \text{diag}(FD_i^T D_i F^*) + \lambda\tau \sum_{i=1}^N \text{diag}(FM_{H_i}^* M_{H_i} F^*) + \tilde{P}) \right). \quad (6)$$

Here $./$ means the componentwise division. One more remark about this approach is that since D_i 's and M_{H_i} 's are circulant matrices which can be diagonalized under the Fourier transform, both $FD_i^T D_i F$ and $FM_{H_i}^* M_{H_i} F^*$ are diagonal matrices. We follow the convention that $0/0 = 0$. The above analysis yields the following algorithm.

Algorithm 1 GeoCS Stage I (solving (5))

1. Initialization: set $u^0, r_i^0, v_i^0, s_i^0, t_i^0$ as zero matrices, and choose proper parameters $\beta, \mu, \lambda, \tau, \gamma > 0$.
2. For $k = 0, 1, 2, \dots$, run the following steps:

$$\begin{aligned} r_i^{k+1} &= \text{shrink}(D_i u^k + v_i^k, 1/\mu), \quad i = 1, 2 \\ s_i^{k+1} &= \text{shrink}(SH_i(u^k) + t_i^k, 1/\tau), \quad i = 1, \dots, N \\ u^{k+1} &\text{ is given by (6)} \\ v_i^{k+1} &= v_i^k + \gamma(D_i u^{k+1} - r_i^{k+1}), \quad i = 1, 2 \\ t_i^{k+1} &= t_i^k + \gamma(SH_i(u^{k+1}) - s_i^{k+1}), \quad i = 1, 2, \dots, N. \\ \text{If } \|u^{k+1} - u^k\| &\leq \text{tol, stop the iteration.} \end{aligned}$$

3.2 Stage II: $w\text{TV-}L_1\text{-}L_2$ model

The stage I model works well in mild CS scenario but not so efficient in challenging scenarios when sampling rate is extremely low and noise is excessive. To handle challenging situations, we start with the result of stage I and then alternatively perform geometric information update and image reconstruction in a beneficial way. Specifically, setting the result from stage I as initial guess, we define adaptive weights based on it, and use weighted TV along with shearlet to reconstruct an image. We then continue alternating weight update and image reconstruction until it converges. For a fixed weight, the model reads as below

$$\min_{u \in \mathbb{R}^{n^2}} \beta \sum_{i=1}^2 \|w_i \cdot D_i u\|_1 + \lambda \sum_{i=1}^N \|SH_i(u)\|_1 + \frac{1}{2} \|\mathcal{F}_p(u) - b\|_2^2 \quad (7)$$

where w_i 's are the weights based on the extracted geometric information, such as reliable gradients and high frequency subbands of shearlet transform coefficients. And \cdot is componentwise multiplication. Algorithm 2 is designed to address the above problem and the complete stage II algorithm is shown in Algorithm 3.

Given the latest iterate \tilde{u} for the reconstructed image, we define TV weight at each pixel as a function of the gradient of \tilde{u} at the same pixel. Suppose $g : [0, \infty) \rightarrow [0, 1]$ is a non-increasing function satisfying

$$g(0) = 1, \quad \lim_{s \rightarrow \infty} g(s) = 0.$$

g is called edge stopping function in image segmentation or diffusivity function in PDE. The reason is that $g(|\nabla \tilde{u}|)$ approaches to zero near edges where the gradient gets large while close to one in smooth areas where the gradient becomes small. In fact, besides separating edges from smooth areas, g also identifies the small differences in intensity variations *within* the smooth areas. The pixel in regions of small intensity variations will get larger g value than that in regions of large intensity variations. Weighted TV with this type of

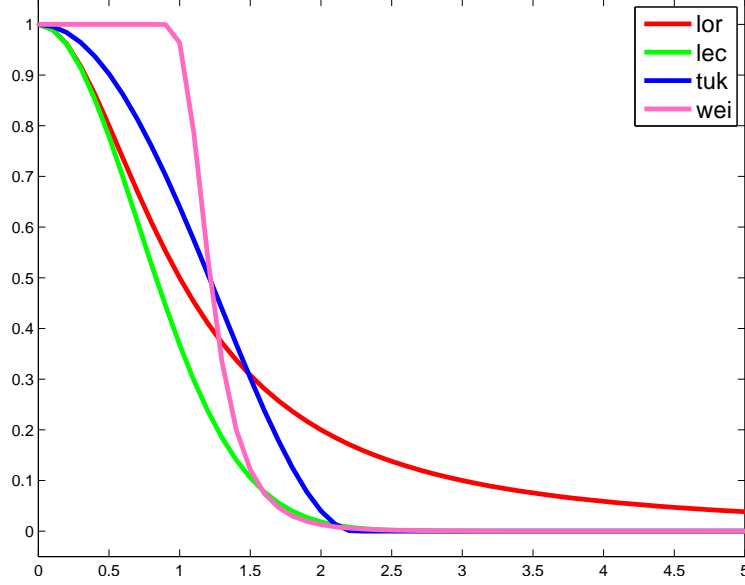


Figure 1: Edge stopping function g 's when $h = 1$

weight will preserve the various intensity variation scales in the reconstruction process, and thus increase the robustness of TV and reduce the staircase effects of TV.

There are many choices for g . Some commonly used ones are listed below, where h is a parameter controlling the differentiation of smoothness levels.

(a) Lorentzian function

$$g_{Lor}(x) = \frac{1}{1 + \frac{x^2}{h^2}}$$

(b) Le Clerc function

$$g_{Lec}(x) = \exp\left(-\frac{x^2}{h^2}\right)$$

(c) Tukey bi-weight function

$$g_{Tuk}(x) = \begin{cases} \left(1 - \frac{x^2}{5h^2}\right)^2 & |x| < \sqrt{5}h \\ 0 & \text{otherwise} \end{cases}$$

(d) Weickert function

$$g_{Wei}(x) = \begin{cases} 1 - \exp\left(-\frac{3.31488h^8}{x^8}\right) & x \neq 0 \\ 1 & \text{otherwise} \end{cases}$$

In Fig. 1, we plot the above four g functions when $h = 1$. From observation, it's clear that they have different decay behaviors. Especially, Weickert edge function decays slowly at the two ends but fast near the middle and Tukey bi-weight function decays slowly all the way long. So for piecewise constant images whose intensity changes sharply from one region to another, Weickert is optimal while Tukey bi-weight function is more appropriate for generic complicated piecewise smooth images with ubiquitous unprecedented intensity variations. We use Tukey bi-weight for all our numerical experiments as we focus on testing piecewise smooth images.

Since we adopt anisotropic TV discretization, our weights are different along x, y directions and are defined respectively as

$$w_1 = g(|D_1 \tilde{u}|), \quad w_2 = g(|D_2 \tilde{u}|). \quad (8)$$

Notice that high frequency components of the shearlet transform of \tilde{u} also provide some edge information. So another option to define weights is to gather all the high frequency subbands. But our massive numerical experiments show that g function of gradients is more efficient.

Algorithm 2 Weighted TV shearlet based image reconstruction algorithm (solving (7))

1. Initialization: set $u^0, r_i^0, v_i^0, s_i^0, t_i^0$ as those generated from stage I.
2. For $k = 0, 1, 2, \dots$, run the steps:

$$\begin{aligned}
r_i^{k+1} &= \text{shrink}(D_i u^k + v_i^k, w_i/\mu), \quad i = 1, 2 \\
s_i^{k+1} &= \text{shrink}(SH_i(u^k) + t_i^k, 1/\tau), \quad i = 1, \dots, N \\
u^{k+1} &\text{ is given by (6)} \\
v_i^{k+1} &= v_i^k + \gamma(D_i u^{k+1} - r_i^{k+1}), \quad i = 1, 2 \\
t_i^{k+1} &= t_i^k + \gamma(SH_i(u^{k+1}) - s_i^{k+1}), \quad i = 1, 2, \dots, N. \\
&\text{If } \|u^{k+1} - u^k\| \leq \text{tol, stop the iteration.}
\end{aligned}$$

In hope of retrieving more trustworthy geometric prior information, we update the weights from each convergent intermediate result and reapply the Algorithm 2. Then we get the Algorithm 3.

Algorithm 3 GeoCS Stage II

1. Initialization: set $u^0, r_i^0, v_i^0, s_i^0, t_i^0$ as those produced from stage I.
 2. For $j = 0, 1, 2, \dots$, run the steps:
 - (1) Build the weights w_1^{j+1}, w_2^{j+1} based on u^j by (8).
 - (2) Set $u^j, r_i^j, v_i^j, s_i^j, t_i^j$ as initial values, apply Algorithm 2 to solve (7) and get $u^{j+1}, r_i^{j+1}, v_i^{j+1}, s_i^{j+1}, t_i^{j+1}$.
 - (3) If $\|u^{j+1} - u^j\| \leq \text{tol}$, stop the iteration.
-

4 Convergence analysis

There are close relationships between Bregman iterative methods and its variants such as linearized Bregman, Bregman operator splitting, and the classical Lagrangian based methods, such as method of multipliers, the alternating direction method of multipliers (ADMM) and alternating minimization algorithm (AMA). The connection between split Bregman algorithm and ADMM, and its illustrative applications in TV- L_1 and TV- L_2 problems can be found in [26]. In this section, we aim to bridge the gap between our proposed algorithms using split Bregman method and ADMM by constructing one augmented Lagrangian adapted to our problem. Then the existing convergence theory for ADMM can be used to justify our proposed algorithm utilizing the split Bregman and quadratic penalties.

We first analyze the algorithm in stage I, and the discussions can be analogously extended to the algorithm 2 in stage II. Based on the problem (4), we build the augmented Lagrangian as below:

$$\begin{aligned}
L(u, r_1, r_2, s_1, \dots, s_N, v_1, v_2, t_1, \dots, t_N) &= \beta \sum_{i=1}^2 (\|r_i\|_1 - v_i^T(r_i - D_i u) + \frac{\mu}{2} \|r_i - D_i u\|_2^2) \\
&\quad + \lambda \sum_{i=1}^N (\|s_i\|_1 - t_i^T(s_i - SH_i(u)) + \frac{\tau}{2} \|s_i - SH_i(u)\|_2^2) + \frac{1}{2} \|\mathcal{F}_p(u) - b\|_2^2.
\end{aligned}$$

Since the variables r_i 's and s_i 's are separable in the Lagrangian L , minimizing L over $(r_1, r_2, s_1, \dots, s_N)$

simultaneously can be replaced by minimizing L over r_i 's and s_i 's individually. Thus ADMM yields

$$\begin{cases} r_i^{k+1} = \underset{r_i}{\operatorname{argmin}} \|r_i\|_1 - (v_i^k)^T(r_i^k - D_i u^k) + \frac{\mu}{2} \|r_i - D_i u^k\|_2^2, & i = 1, 2 \\ s_i^{k+1} = \underset{s_i}{\operatorname{argmin}} \|s_i\|_1 - (t_i^k)^T(s_i - SH_i(u^k)) + \frac{\tau}{2} \|s_i - SH_i(u^k)\|_2^2, & i = 1, \dots, N \\ u^{k+1} = \underset{u}{\operatorname{argmin}} L(u, r_1^{k+1}, r_2^{k+1}, s_1^{k+1}, \dots, s_N^{k+1}, v_1^k, v_2^k, t_1^k, \dots, t_N^k) \\ v_i^{k+1} = v_i^k + \mu\gamma(D_i u^{k+1} - r_i^{k+1}), & i = 1, 2 \\ t_i^{k+1} = t_i^k + \tau\gamma(SH_i(u^k) - s_i^{k+1}), & i = 1, \dots, N \end{cases}$$

By absorbing the linear terms involving v_i 's and t_i 's into the quadratic terms, it can be simplified as

$$\begin{cases} r_i^{k+1} = \underset{r_i}{\operatorname{argmin}} \|r_i\|_1 + \frac{\mu}{2} \left\| r_i - D_i u^k - \frac{v_i^k}{\mu} \right\|_2^2, & i = 1, 2 \\ s_i^{k+1} = \underset{s_i}{\operatorname{argmin}} \|s_i\|_1 + \frac{\tau}{2} \left\| s_i - SH_i(u^k) - \frac{t_i^k}{\tau} \right\|_2^2, & i = 1, \dots, N \\ u^{k+1} = \underset{u}{\operatorname{argmin}} \frac{\beta\mu}{2} \sum_{i=1}^2 \|r_i^{k+1} - D_i u\|_2^2 + \frac{1}{2} \|\mathcal{F}_p(u) - b\|_2^2 + \frac{\lambda\tau}{2} \sum_{i=1}^N \|s_i^{k+1} - SH_i(u)\|_2^2 \\ v_i^{k+1} = v_i^k + \mu\gamma(D_i u^{k+1} - r_i^{k+1}), & i = 1, 2 \\ t_i^{k+1} = t_i^k + \tau\gamma(SH_i(u^k) - s_i^{k+1}), & i = 1, \dots, N \end{cases} \quad (9)$$

One can see the algorithm derived by ADMM here is equivalent to Algorithm 1 by split Bregman method with quadratic penalization. The detailed convergence analysis of the algorithm (9) can be found in [26]. To be complete, we present the convergence theorem without the proof.

Theorem 1. *For any $\mu, \tau > 0$ and $\tilde{\gamma} \in (0, (\sqrt{5} + 1)/2)$, the sequences $\{(u^k, r_i^k, s_i^k)\}$ generated by (9) from any starting point (u^0, λ^0, η^0) converges to a solution of problem (5).*

Therefore, by choosing appropriate parameter γ , the proposed Algorithm 1 provides a convergent solution to the problem (4). Likewise for the fixed weights w_i 's, by replacing $D_i u^k$ with $w_i * D_i u^k$ the ADMM yields a similar algorithm equivalent to Algorithm 2 and thereby the convergence is guaranteed as well.

Parameter selection. The above theorem only requires $\gamma \in (0, (\sqrt{5} + 1)/2)$, and positive μ, τ to guarantee convergence. In the perspective of convergence speed, our experience with a variety of tests shows that γ restricted in $[1, (\sqrt{5} + 1)/2]$ consistently yields good results. Regarding μ and τ , as they show up in the shrinkage representation of updates for r_i and s_i , an inappropriate selection of them leads to slow convergence. Especially, if μ and τ are set too small, the updates for r_i and s_i will dwell in 0 at the first several iterations. We scale image intensity to $[0, 1]$ to make the effect of μ and τ on convergence speed moderate. β and λ depend on the gradient/shearlet transform sparsity of the underlying image and the noise/error level in the measurements. Implementation details and specific parameter selections will be explained in Section 5.

5 Numerical examples

In this section, we illustrate the performance of GeoCS on various images with different sampling rates and noise levels. We also compare GeoCS with two recent related CS reconstruction approaches: RecPF [17] and EdgeCS [15]. All experiments were performed under Windows 7 Professional operating system and MATLAB R2012a running on a Dell desktop with Intel Core i5 CPU at 3.10 GHz and 8 GB of memory.

RecPF iteratively recovers an image from its incomplete Fourier samples by solving

$$\min_u \beta \text{TV}(u) + \lambda \|\Phi u\|_1 + \frac{1}{2} \|\mathcal{F}_p(u) - b\|_2^2 \quad (10)$$

where $\text{TV}(u)$ can be either isotropic or anisotropic and Φ is wavelet transform.

Sampling rate	8.79%	10.64%	12.92%	14.74%
Proposed	0.1200	0.1016	0.0874	0.0797
RecPF	0.1468	0.1273	0.1097	0.1011
EdgeCS	0.1581	0.1378	0.1208	0.1111

Table 1: Relative error comparison for brain MRI reconstruction

EdgeCS alternatively performs image reconstruction and edge detection in a mutually beneficial manner. It detects edges from the intermediate reconstruction and use edge information to guide the next stage of image reconstruction and so on. GeoCS is different from EdgeCS as analyzed in Section 1.1.

Our test images are all piecewise smooth images with a lot of fine details: a human brain MR image, Barbara image with textures and a human knee MR image. The intensity value of each test image is scaled to the range $[0, 1]$ before simulating b . Partial Fourier CS data are simulated through fast Fourier transform (FFT) on the test images followed by sampling on smooth radial trajectories that are empirically shown to be effective.

All the quantitative comparisons are based on relative error and signal-to-noise ratio (SNR). Relative error is to measure the recovery accuracy and defined as

$$\text{Relerr} = \frac{\|u - u_{\text{true}}\|_2^2}{\|u_{\text{true}}\|_2^2}$$

where u and u_{true} are the recovered image and the ground truth, respectively. Considering the independence with the above measure, we adopt the SNR defined in [27]

$$\text{SNR} = 10 \log_{10} \frac{\|u.^2 + u_{\text{true}}.^2\|_2^2}{\|u - u_{\text{true}}\|_2^2},$$

where $(.^2)$ represents the componentwise squaring. For all the experiments, we fix $\gamma = 1, \mu = \tau = 10^2$ and vary β, λ slightly based on the noise level. When there is no noise we set $\beta = \lambda = 10^{-5}$ for all of the three images while they are set a little larger in the presence of noise. The results are not sensitive to the selection of β and λ . For discrete shearlet transform, we adopt FFST [21] with 3 scales and 13 subbands (12 high frequency and one low frequency). The parameter h used in Tukey bi-weight g function in stage II is set among $(0, 1]$. And in each of the following experiment, the total number of iterations used in stage I is less than 1000 and stage II takes less than 100 iterations to achieve a convergent solution.

5.1 Example 1

In the first example, we look at the simulated noise-free spectral measurements of a 512×512 brain MR image. The ground truth image has inhomogenous contrasts in different areas, especially in the gray matter and cerebrospinal fluid. We tested the proposed GeoCS algorithm, RecPF and EdgeCS with 40 radial sampling lines, namely 8.79% sampling rate. We show the results in Fig. 2 and zoom in one small patch for better visual comparison. It's apparent that the image produced by GeoCS has better quality than the others. RecPF sort of oversmooths the whole image, and EdgeCS is able to detect the edges while losing some gradual transition between smooth areas and boundaries. To further compare three results, we take the difference between the ground truth and the reconstructed image for each method and display the inverted residue images in Fig. 3. It's clear that our proposed algorithm suppresses the error more evenly inside the skull. The three approaches are also compared as the sampling rate changes. The quantitative comparison listed in Table 1 and Table 2 shows that the proposed GeoCS consistently outperforms the other methods.

5.2 Example 2

In this example, we show the benefits of GeoCS on reconstructing texture images. The test image is 512×512 Barbara image which has various texture patterns and a lot of details. This image requires relatively higher sampling rate to get an ideal recovery and standard edge detection algorithm may even fail to get accurate

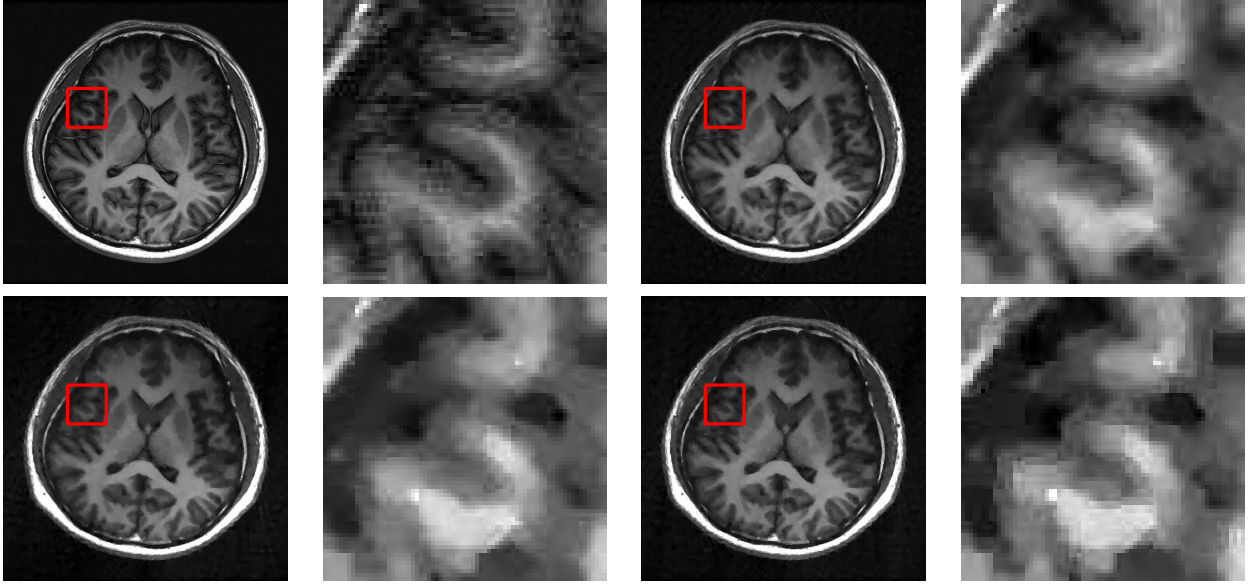


Figure 2: Reconstructed brain MR image. First row from left to right: ground truth, close-up of ground truth, our result, close-up of the result. Second row from left to right: result obtained by RecPF, close-up of RecPF result, result by EdgeCS, close-up of EdgeCS result.

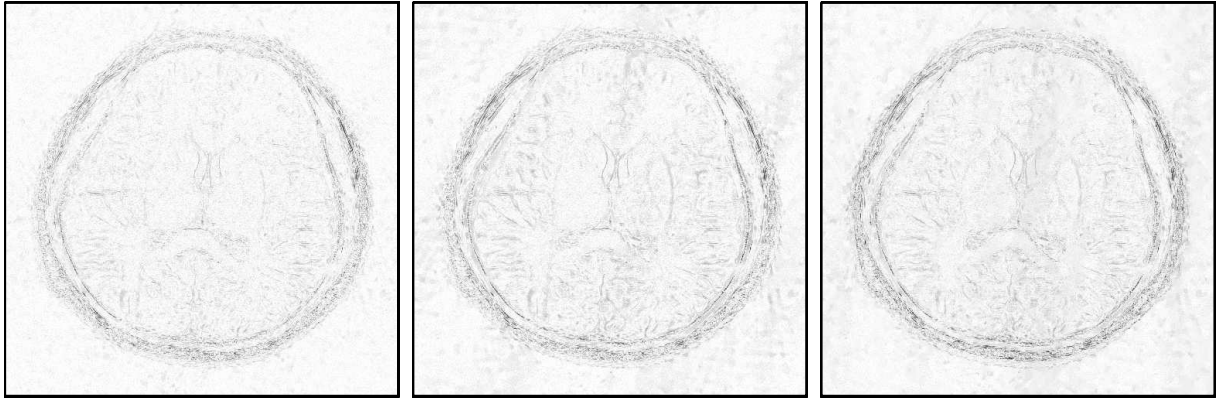


Figure 3: Residual maps of the three method results with the ground truth. For better visualization, we inverted the grayscale. From left to right: proposed, RecPF, EdgeCS, the relative errors are listed respectively as: 12.00%, 14.68%, 15.81%.

Sampling rate	8.79%	10.64%	12.92%	14.74%
Proposed	15.90	17.34	18.65	19.46
RecPF	14.16	15.40	16.68	17.39
EdgeCS	14.06	15.34	16.56	17.35

Table 2: SNR comparison for brain MRI reconstruction

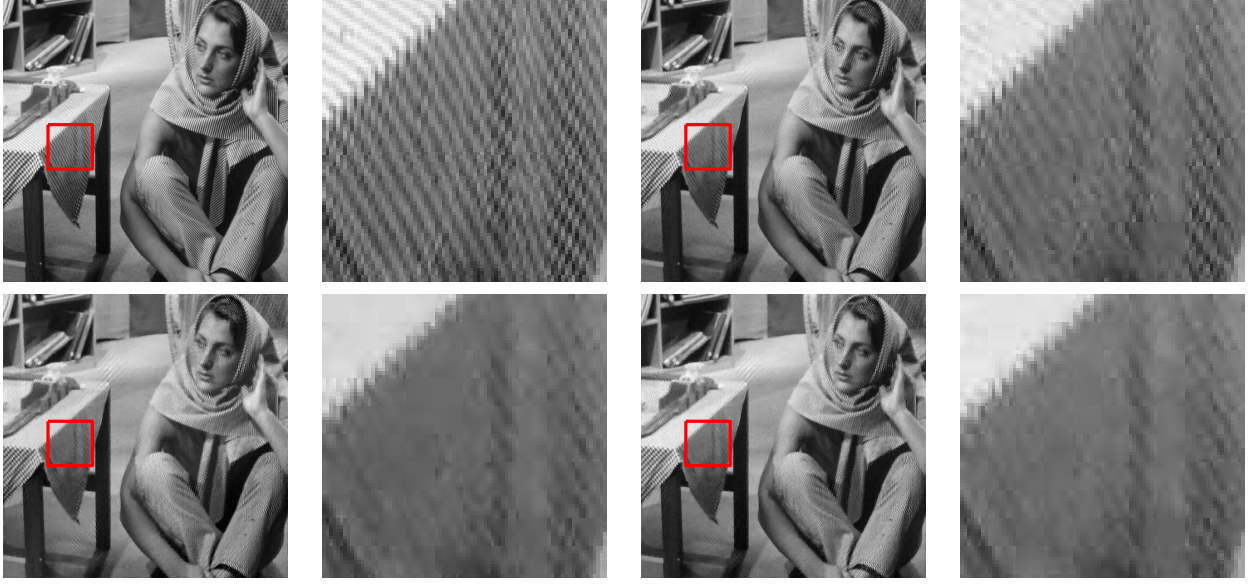


Figure 4: Reconstructed barbara image. First row from left to right: accurate barbara image, cropped barbara image, our result, close-up of our result. Second row from left to right: result by RecPF, close-up of RecPF result, result by EdgeCS, close-up of EdgeCS result.

edges. The results obtained by GeoCS, RecPF and EdgeCS with 100 radial sampling lines (sampling rate 20.87%) are listed in Fig. 4, where we zoomed in one patch of table cloth. Our proposed method is able to recover largely the directional textures while the other two methods get blurry textures. The inverted residue images are listed in Fig. 5. The consistent performance is illustrated in Table 3 and Table 4 using different sampling rates. In this example, shearlet transform plays an important role in preserving the structures in different directions and thereby the textures. Weighted total variation further corrects the smooth areas which were over-texturized.

5.3 Example 3

Our last test image is a T1 weighted MR image of the knee showing femur, patella, tibia and menisci from <http://www.mr-tip.com/>. We first added zero-mean complex Gaussian noise $\sigma = 10$ to the spectral data

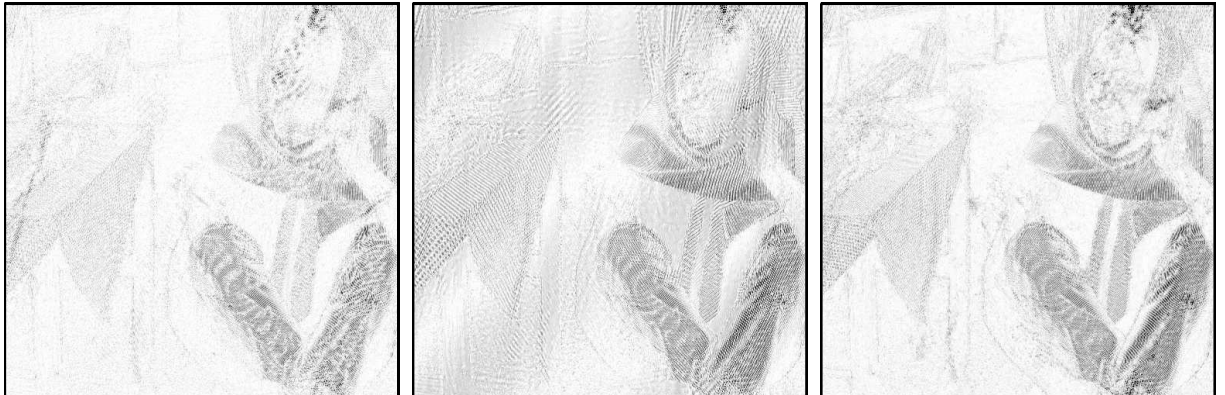


Figure 5: From left to right: proposed, RecPF, EdgeCS. Relative error: 9.64%, 14.01%, 12.53%.

Sampling rate	8.79%	12.92%	16.94%	20.87%
Proposed	0.1429	0.1248	0.1104	0.0964
RecPF	0.1700	0.1552	0.1471	0.1401
EdgeCS	0.1574	0.1412	0.1323	0.1253

Table 3: Relative error comparison for barbara image reconstruction

Sampling rate	8.79%	12.92%	16.94%	20.87%
Proposed	10.16	11.33	12.39	13.58
RecPF	8.66	9.45	9.92	10.36
EdgeCS	9.69	10.72	11.35	11.92

Table 4: SNR comparison for barbara image reconstruction

sampled by 40 radial lines (sampling rate 12.71%). The recovered images and their associated enlarged patches given by GeoCS, RecPF and EdgeCS are shown respectively in Fig. 6. Visually our result is more natural in the bones and junctions and is closer to the ground truth than those given by the other two methods. Again for better visual comparison, the residue images are listed in Fig. 7.

By fixing the sampling rate as 12.71%, we also perform the comparisons on spectral data with different noise levels $\sigma = 5, 10, 15, 20$. In the case of large noise level, it's better to adjust the regularization parameters accordingly. It's also true in RecPF and EdgeCS algorithms. From Table 5 and Table 6, we observe that our proposed algorithm is robust to the noise and produces more accurate reconstructed images than the other methods. All the results are obtained under the optimal parameter settings for each method.

6 Conclusion and remarks

We proposed a two-stage compressive sensing image reconstruction algorithm based on shearlet transform and weighted TV. The first stage is to use standard TV- L_1 - L_2 model with shearlet transform to get an initial guess for the underlying image of interest. Geometric information extracted from this guess serves as an initial a priori in weighted TV- L_1 - L_2 model to further enhance the reconstruction accuracy. This kind of geometric information extraction and image reconstruction are alternated in a mutually beneficial fashion until it converges. Replacing the conventional wavelet transform with shearlet transform, the model is able to promote the signal sparsity, and preserve multiple directional features better during the recovery. The spatially variant weights associated to TV plays an important part in preserving sharp edges while reducing staircase effects of TV. The minimization problem is solved by split Bregman which divides one complicated optimization problem with nondifferentiable terms into three sets of subproblem, each of which has closed-form solutions. Convergence of the algorithm is guaranteed under mild conditions. The proposed approach is compared with two recent related work. Numerical experiments show the consistent overwhelming advantages of our algorithm.

The proposed approach GeoCS is better than RecPF and EdgeCS in reconstructing complicated piecewise smooth images. However, as for piecewise constant images, it's sufficient to apply one-stage methods. Moreover, by adapting the weights to the spatially variant gradients along with two-stage reweighting scheme, GeoCS integrates more reliable geometric prior to the reconstruction than EdgeCS. Our extensive experience shows that the more accurate geometric information is obtained during the algorithm, the better the overall scheme will perform. But, it is possible that the extracted geometric information is not reliable at all in

σ	5	10	15	20
Proposed	0.0842	0.0952	0.1006	0.1095
RecPF	0.1064	0.1171	0.1283	0.1389
EdgeCS	0.0992	0.1035	0.1118	0.1194

Table 5: Relative error comparison for noisy knee MRI reconstruction

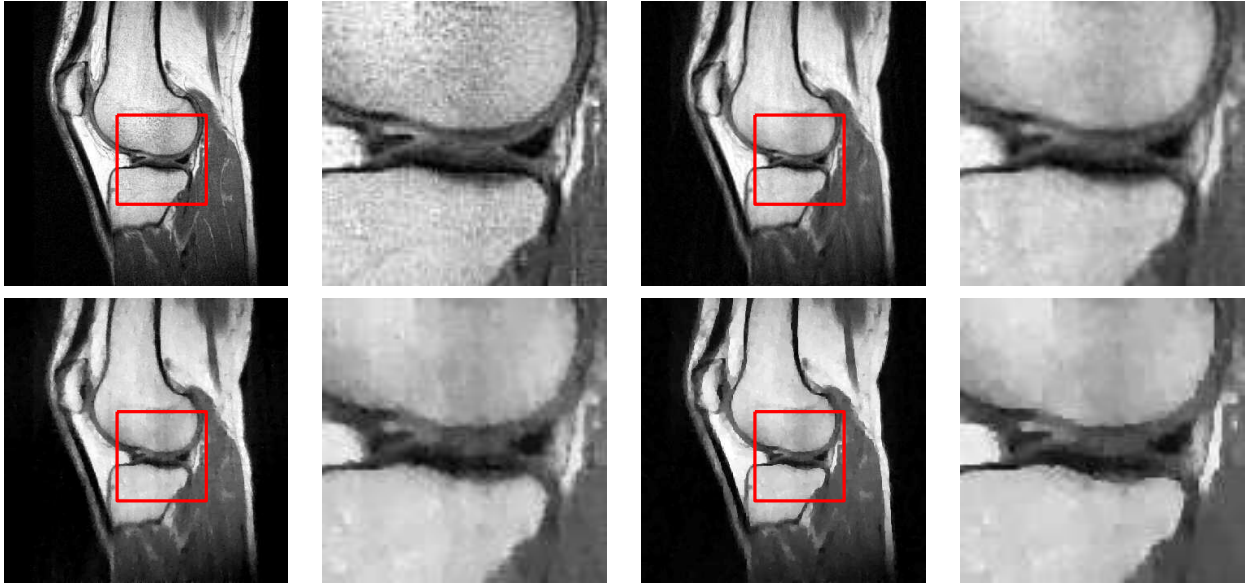


Figure 6: Recovered knee image from noisy data. Top row from left to right: ground truth, our result. Bottom row from left to right: result by RecPF and EdgeCS.

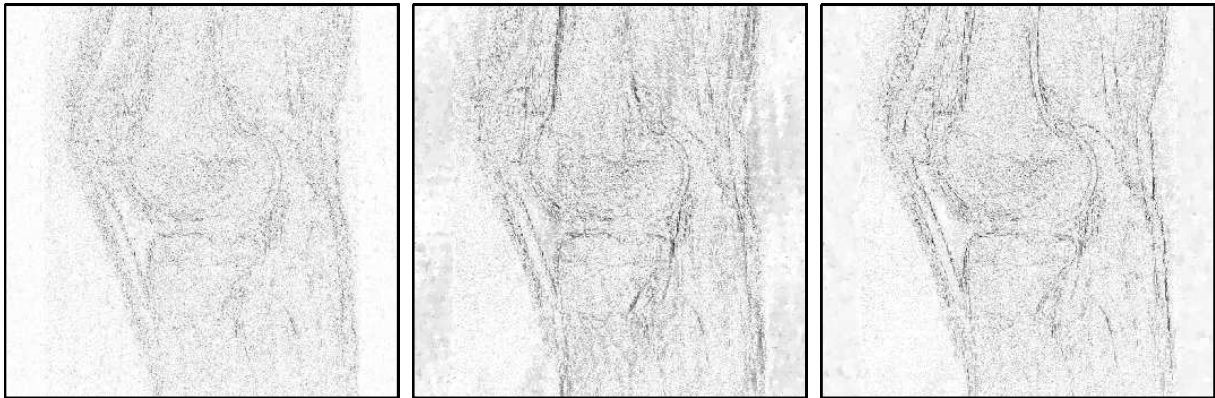


Figure 7: From left to right: proposed, RecPF, EdgeCS. Relative error: 9.52%, 11.71%, 10.35%.

σ	5	10	15	20
Proposed	18.50	17.44	16.95	16.22
RecPF	16.47	15.64	14.85	14.16
EdgeCS	17.08	16.71	16.04	15.47

Table 6: SNR comparison for noisy knee MRI reconstruction

case of extremely insufficient samples or excessive noise and thereby GeoCS might fail. There is still room to study how to extract much more reliable geometric information from noisy incomplete measurements and how to efficiently utilize them. Furthermore, some recent acceleration techniques, e.g., Nesterov's accelerated gradient descent[28], can be applied to speed up the convergence.

Acknowledgement

The authors would like to thank the partial support from the Case Western Reserve University ACES+ advance opportunity funds.

References

- [1] Candès, E., Tao, T.: Near optimal signal recovery from random projections: universal encoding strategies. *IEEE Transactions on Information Theory*. 52(1), 5406-5425 (2006)
- [2] Donoho, D.: Compressed sensing. *IEEE Transactions on Information Theory*. 52, 1289-1306 (2006)
- [3] Candès, E., Eldar, Y., Needell, D., Randall, P.: Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1), 59-73 (2011)
- [4] Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52, 489-509 (2006)
- [5] Lustig, M., Donoho, D., Pauly, J.: Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6), 1182-95 (2007)
- [6] Compton, R., Osher, S., Bouchard, L.: Hybrid regularization for MRI reconstruction with static field inhomogeneity correction. *UCLA CAM-Reports* 11-84 (2011)
- [7] Cai, J.F., Dong, B., Osher, S., Shen, Z.W.: Image restoration: Total variation, wavelet frames, and beyond. *J. Amer. Math. Soc.* 25, 1033-1089 (2012)
- [8] Labate, D., Lim, W.-Q., Kutyniok, G., Weiss, G.: Sparse multidimensional representation using shearlets. *Proc. SPIE Wavelets XI*, San Diego, CA, 5914, 254-262 (2005)
- [9] Kutyniok, G., Guo, K., Labate, D.: Sparse multidimensional representations using anisotropic dilation and shear operators. *Wavelets and Splines* (Athens, GA, 2005), Nashboro Press, Nashville, TN (2006)
- [10] Easley, G., Labate, D., Lim, W.-Q.: Sparse directional image representation using the discrete shearlet transform. *Appl. Comput. Harmon. Anal.*, 25, 25-46 (2008)
- [11] Goldstein, T., Osher, S.: The Split Bregman Method for L1 Regularized Problems. *SIAM Journal on Imaging Sciences*, 2(2): 323-343 (2009)
- [12] Glowinski, R., Marroco, A.: Sur L'Approximation, par Elements Finis d'Ordre Un, et la Resolution, par Penalisation-Dualité, d'une Classe de Problems de Dirichlet non Lineares. *Revue Française d'Automatique, Informatique et Recherche Opérationnelle* 9(R-2), 41-76 (1975).
- [13] Gabay, D., Mercier B.: A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximations. *Computers and Mathematics with Applications* 2, 17-40 (1976).
- [14] Wang, Y., Yin, W.: Sparse signal reconstruction via iterative support detection. *SIAM Journal on Imaging Sciences*, 3(3), 462-491 (2010)
- [15] Guo, W., Yin, W.: EdgeCS: Edge guided compressive sensing reconstruction. *Proceedings of SPIE Visual Communication and Image Processing*, 7744, 77440L1-L10 (2010)
- [16] Guo, W., Yin, W.: Edge guided reconstruction for compressive imaging. *SIAM J. Imaging Sci.*, 5(3), 809-834 (2012)

- [17] Yang, J., Zhang, Y., Yin, W.: A fast alternating direction method for TVL1-L2 signal reconstruction from partial fourier data. *Selected Topics in Signal Processing, IEEE Journal of*, 4(2), 288-297 (2010)
- [18] Candès, E., Donoho, D.: Ridgelets: a key to higher-dimensional intermittency? *Phil. Trans. R. Soc. Lon. A.*, 357(1760), 2495-2509 (1999)
- [19] Candès, E., Donoho, D.: Curvelets - a surprisingly effective nonadaptive representation for objects with edges. *Curves and Surfaces*, L. L. Schumaker et al. (eds), Vanderbilt University Press, Nashville, TN.
- [20] Do, M. N., Vetterli, M.: The contourlet transform: An efficient directional multiresolution image representation. *IEEE Trans. Imag. Proc.*, 14(12), 2091-2106 (2005)
- [21] Häuser, S.: Fast finite shearlet transform: a tutorial. *ArXiv* 1202.1773 (2012)
- [22] Easley, G., Labate, Colonna, F.: Shearlet-based total variation diffusion for denoising. *Image Processing, IEEE Transactions on*, 18(2), 260-268 (2009)
- [23] Wang, F., Wang, S., Hu, X., Deng, C.: Compressive Sensing of Image Reconstruction Based on Shearlet Transform. *Advances in Intelligent and Soft Computing*, 125, 445-451 (2012)
- [24] Hale, E., Yin, W. and Zhang, Y.: A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing. *CAAM Technical Report-TR07* (2007)
- [25] Pizurica, A.: A Bayesian formulation of edge-stopping functions in nonlinear diffusion. *IEEE Signal Processing Letters*, 13(8), 501-504 (2006)
- [26] Esser, E.: Applications of Lagrangian-Based Alternating Direction Methods and Connections to Split Bregman. *UCLA CAM-reports*, 09-31 (2009)
- [27] Coupé, P., Hellier, P., Kervrann, C., Barillot, C.: Bayesian non local means-based speckle filtering. *Proceedings of ISBI*, 1291-1294 (2008)
- [28] Nesterov, Y.: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Doklady AN SSSR* (translated as *Soviet Mathematics Doklady*), 269, 543-547 (1983)